

A Principal Component Analysis of TRACE-P Whole Air Data: Halocarbons, Hydrocarbons, Alkyl Nitrates and Sulfur Compounds

Isobel J. Simpson,¹ Yunsoo Choi,^{1,2} Donald R. Blake,¹ Simone Meinardi,¹ Nicola J. Blake,¹ and F. Sherwood Rowland¹

¹ Department of Chemistry, University of California-Irvine, Irvine, California, USA

² Now at Georgia Institute of Technology, Atlanta, Georgia, USA

Abstract. Principal Component Analysis (PCA) is a multivariate statistical technique that can be used to help identify major pollution sources and their emission composition from within a large ambient measurement data set. Here we apply PCA to a regional whole air data set collected during the airborne Transport and Chemical Evolution over the Pacific (TRACE-P) field campaign (February-April, 2001). The data set comprises a matrix of 3458 samples by 47 compounds (halocarbons, nonmethane hydrocarbons, alkyl nitrates, and sulfur compounds), collected at altitudes between 0.1-12 km and west of 165°W. Knowledge of trace gas emission signatures was used to interpret the PCA results. Whereas complex source signatures were sometimes observed north of 25°N in the free troposphere (2-7 km) and south of 25°N between 0.1-12 km, most of the derived principal components were explained by one or two emission sources. Between 25-46°N, industrial tracers such as the CFC replacement compounds (HCFCs) dominated the variance within the data set. Photochemically processed urban air mixed with oceanic tracers, and air influenced by emissions from the People's Republic of China also featured strongly. Further south between 7-25°N, processed urban air dominated the observed variance. Tracers indicative of automobile exhaust and gasoline evaporation were also evident,

as were oceanic signatures and industrial influences including the CFC family. The variability of CFC mixing ratios detected using PCA reflects their continuing use within Asia. Overall, PCA served as a compact, robust method for simplifying the large, regional TRACE-P data set, and for identifying specific emission sources and regions.

1. Introduction

Pollutant emissions from Asia impact the atmosphere on local to intercontinental scales, and are an important contributor to global change [*IPCC*, 2001a; *Jacob et al.*, 2003 and references therein]. Whereas the emissions of greenhouse gases from China (e.g. carbon dioxide and methane) have declined in recent years [*Streets et al.*, 2001], other compounds related to energy use (e.g. nitrogen oxides and nonmethane volatile organic carbons) have shown a significant increase [*Streets et al.*, 2003]. Currently, the strongest growth rates in global energy use occur in developing countries of the Asia-Pacific region [*IPCC*, 2001b].

The airborne Transport and Chemical Evolution over the Pacific (TRACE-P) experiment was conducted in spring, 2001 to measure the chemical composition and evolution of outflow from Asia to the Pacific, and to quantitatively relate the pollution to its sources [*Jacob et al.*, 2003]. In support of these objectives, more than 3400 whole air samples were collected over the western Pacific between altitudes of 0.1-12 km. Each whole air sample was analyzed for dozens of compounds comprising halocarbons, nonmethane hydrocarbons, alkyl nitrates, and sulfur compounds. The resulting data matrix includes more than 160,000 data points, and it is a size for which multivariate statistical techniques become valuable as analytical tools. Principal Component Analysis (PCA) is used to help identify major pollution sources and their emission composition from within large ambient measurement data sets. Here we apply PCA to the TRACE-P whole air data set to identify its major modes of composition and to determine which

components drive its variability. Trace gas emission signatures are then used to assign pollution sources to the derived principal components.

2. Experimental

The TRACE-P field campaign took place over the western and central Pacific Ocean in the spring of 2001 (February 24-April 10). Whole air samples were collected aboard the NASA DC-8 and P-3B aircraft during 38 science flights of the TRACE-P mission, which had major deployment sites in Hong Kong (22°N, 114°E) and the Yokota Air Force Base in Fussa, a Tokyo suburb (36°N, 140°E). The DC-8 can achieve a maximum altitude of 12 km, whereas the P-3B is designed for lower altitudes and reaches a maximum height of 7 km. A detailed overview of the sampling area, including flight tracks for both aircraft, is given in *Jacob et al.* [2003].

During each flight, individual whole air samples were collected in 2-L stainless steel canisters, each equipped with a stainless steel bellows valve. Before being brought into the field, the canisters were conditioned and evacuated, and 20 Torr of distilled, degassed water was added into each canister in order to quench active surface sites. Air samples were collected roughly every 3-7 minutes during horizontal flight legs, and every 1-3 minutes during ascents and descents. The air flow rate was regulated so that it took about one minute to fill each canister, regardless of the external atmospheric pressure. Approximately 160 samples were collected per flight aboard the DC-8, and 140 aboard the P-3B. In this paper, air samples collected between 7-46°N; east of 112°E and west of 165°W; and between altitudes of 0.1-12 km were used (3458 samples). We note that take-off and landing data collected below 100 m were not used in our analysis.

After each flight, the whole air samples were returned to our laboratory at the University of California, Irvine (UC-Irvine). The samples were analyzed for 78 trace gases within 10 days

of being collected, using gas chromatography (GC) coupled with flame ionization detection (FID) and mass spectrometer detection (MSD) for nonmethane hydrocarbons (NMHCs); GC with MSD and electron capture detection (ECD) for halocarbons and alkyl nitrates; and GC with MSD for sulfur compounds. For the work presented in this paper, we applied PCA to a subset of 47 compounds (26 C₁-C₂ halocarbons, 11 C₂-C₇ nonmethane hydrocarbons, 7 C₁-C₅ alkyl nitrates, and 3 C₁-C₂ sulfur compounds). The 47 individual species are listed in Table 1, and these are the compounds that were routinely present at mixing ratios greater than their limit of detection (LOD) during TRACE-P. They represent a diverse range of emission signatures (Table 2; Section 3.2) and atmospheric lifetimes (1 day – 300 years), such that indicators of numerous sources and atmospheric processing times are represented in the TRACE-P whole air data set.

The analysis of whole air samples at UC-Irvine, including calibration procedures, is fully described in *Colman et al.* [2001]. Additional analytical information specific to the TRACE-P mission is found in *Blake et al.* [2003a; this issue] and *Simpson et al.* [2003]. Briefly, for each sample, $1520 \pm 1 \text{ cm}^3$ (STP) of air was pre-concentrated on glass beads in a liquid nitrogen-cooled stainless steel loop. Hot water was used to revolatilize the components, which were then flushed to a splitter that partitioned the flow into five different streams. Each stream was sent to one of five column-detector combinations (DB1/FID; PLOT-DB1/FID; Restek-1701/ECD; DB5-Restek1701/ECD and DB5/MSD). The addition of 20 Torr of water into each canister ensures that the split ratios (15.1, 60.8, 7.2, 6.8 and 10.1%, respectively) are highly reproducible.

The accuracy of the measurements varies by compound, and ranges from 2-20% for the trace gases reported here. The measurement precision varies both by compound and by mixing ratio. For example, the measurement precision is 1% or 1.5 pptv (whichever is larger) for the alkanes and alkynes, and 3% or 3 pptv (whichever is larger) for the alkenes. By comparison, the

precision for C₂Cl₄ at a mixing ratio of 5 pptv is ± 0.05 pptv. The LOD is 3 pptv for the NMHCs. The reader is referred to *Colman et al.* [2001] for a complete listing of accuracy, precision, and LOD for each reported species.

3. Data Analysis and Interpretation

3.1 Principal Component Analysis

Receptor models are used to relate concentrations measured at a given location (the receptor site) backward to the responsible emission sources, without requiring a knowledge of the dispersion pattern of the pollutants [*Seinfeld and Pandis*, 1998]. Principal Component Analysis is a multivariate receptor modeling technique that replaces a large number of intercorrelated variables with a smaller set of independent variables [*Thurston and Spengler*, 1985]. In the case of our whole air samples, each new variable (principal component or PC) represents either one pollution source, or in some cases a mixed pollution source. Natural sources such as oceans may also be represented in a principal component. Our objective in applying PCA is to derive a small number of PCs, or emission sources, that explain a maximum of variance in the data. Here we briefly review the basis of Principal Component Analysis. The reader is referred to *Henry and Hidy* [1979], *Thurston and Spengler* [1985] and *Seinfeld and Pandis* [1998] for detailed descriptions and examples of PCA.

The TRACE-P whole air data set collected over the western Pacific comprises a matrix of $k = 1 \dots 3458$ rows (the number of whole air samples) by $i = 1 \dots 47$ columns (the number of volatile organic carbon species). The first step in PCA is to transform this data matrix into a dimensionless form, Z_{ik} [*Thurston and Spengler*, 1985]:

$$Z_{ik} = \frac{(C_{ik} - \bar{C}_i)}{\sigma_i} \quad (1)$$

where C_{ik} is the measured concentration of species i for sample k ; \overline{C}_i is the mean concentration of species i ; and σ_i is the standard deviation of species i . In order for PCA to produce stable results, the number of samples k must greatly exceed the number of measured species i . Whereas *Thurston and Spengler* [1985] suggest a rough guide of $k \geq i + 50$, we have found that stable results can be sometimes achieved using a smaller sample number excess (e.g. $k \geq i + 25$) [*Choi et al.*, 2003]. However for the present work, a very large excess of samples to species ($k \geq i + 200$) was always used when subsets of the TRACE-P data set were analyzed.

The second step in PCA is to create an unrotated principal component model [*Yu and Chang*, 2000]. The basic PC model expresses the total concentration of species i as the sum of individual contributions from $j = 1 \dots n$ pollution sources:

$$Z_{ik} = \sum_{j=1}^n A_{ij} P_{jk} \quad (2)$$

where A_{ij} is the source emission signature (the fraction of species i in the emission source j) and P_{jk} is the source contribution (the concentration of pollutants from source j measured in sample k).

Next, we invert the basic unrotated PC model to derive the PC scoring matrix $[B]$ [*Henry and Hidy*, 1979; *Thurston and Spengler*, 1985]:

$$P_{jk} = \sum_{i=1}^n \frac{A_{ji}}{\lambda_j} Z_{ik} = [B]_{j \times i} [Z]_{i \times k} \quad (3)$$

where λ_j is the eigenvalue associated with P_j . As described by *Thurston and Spengler* [1985], “[B] is derived so that the first principal component (PC1) explains as large a per cent of the original variables’ total variance as possible. The coefficients for the second principal component (PC2) are, in turn, chosen so that it explains as large a per cent of the remaining variance in the

original variables (i.e. not explained by PC1), subject to the restriction that PC1 and PC2 are uncorrelated.” These PC coefficients are derived from the correlation matrix $[R]$:

$$[R]_{i \times i} = [Z]_{i \times i} [Z]_{i \times i}^t \quad (4)$$

where the superscript t indicates the transpose of matrix $[Z]_{i \times i}$.

From matrix algebra, if the $i \times i$ matrix $[R]$ has i linearly independent eigenvectors, then it is diagonalizable [Edwards and Penney, 1988]. Therefore, the fourth step in PCA is to diagonalize the correlation matrix, in order to find the uncorrelated components. This is achieved by determining the eigenvectors and eigenvalues of the correlation matrix [Henry and Hidy, 1979; Thurston and Spengler, 1985]:

$$[Q^{-1}]_{i \times i} [R]_{i \times i} [Q]_{i \times i} = [\Lambda]_{i \times i} \quad (5)$$

where $[Q]$ is composed of the eigenvectors that diagonalize the correlation matrix, and $[\Lambda]$ is a diagonal matrix of eigenvalues, arranged in descending order. The principal component associated with the largest eigenvalue is PC1, the second eigenvalue is associated with PC2, and so forth [Henry and Hidy, 1979]. The variance in the data set that is explained by each PC (or linear combination of variables) is proportional to its eigenvalue. Here we retain only those PCs that have eigenvalues greater than 1.0, or in other words, the PCs should explain more of the variance than do individual species [Van de Geer, 1971].

In the final step, a Varimax rotation is applied in order to make the principal components more easily interpreted [Kaiser, 1958]. Without this rotation, it is difficult to associate individual PCs with specific pollution sources [Thurston and Spengler, 1985; Yu and Chang, 2000].

The results of the Principal Component Analysis with Varimax rotation are presented as loading tables (see Tables 3-5). A ‘loading’ is a number between +1 and -1, where a positive loading is a correlation and a negative loading is an anticorrelation. Only positive loadings are

reported and discussed here. In terms of loading order, a higher loading associated with a given species i indicates its greater contribution to the observed variance in the data set.

We have performed several tests to investigate the limitations of the PCA technique. The reader is referred to *Choi et al.* [2003] for additional discussion of the advantages and drawbacks of PCA. For the TRACE-P study, we performed PCA on subsets of data that were separated by latitude (north and south of 25°N) and by altitude (0.1-2 km; 2-7 km; 7-12 km) (Section 4). However, a further division of the data by longitude (east and west of 140°E) yielded comparable loading tables (see Section 4.4 for selected results). In this study we also found that our results are consistent with findings based on other techniques such as meteorological backward trajectories, bivariate analyses and a knowledge of active source regions (e.g. biomass burning areas) during TRACE-P (Sections 4.1 and 4.2). Finally, although PCA has been applied to air pollution data sets for over two decades [*Henry and Hidy*, 1979; *Thurston and Spengler*, 1985; *Statheropoulos et al.*, 1998; *Yu and Chang*, 2000; *Choi et al.*, 2003], newer receptor modeling techniques such as the Positive Matrix Factorization (PMF) method have also been developed [*Paatero*, 1997]. Principal Component Analysis with Varimax rotation (the same program that was used in this study) was found to yield comparable results to the PMF method when both techniques were applied to the regional Tropospheric Ozone Production about the Spring Equinox (TOPSE) data set [*Y. Wang et al.*, 2003]. In general we view PCA as a complement to more familiar methods such as bivariate plots, and it provides a concise overview of correlating species within a large, regional data set.

3.2 Emission Signatures

As described above, PCA is used to transform a large number of intercorrelated variables into a smaller number of independent variables, where each independent variable represents either one

emission source or else a mixed emission source. In order to link these new independent variables to individual pollution sources, a prior knowledge of relevant trace gas emission signatures is required. For example, light (C_2 - C_5) NMHC emissions are linked to pollution sources including natural gas use (e.g. ethane, propane), fossil fuel combustion (e.g. ethene, ethyne, *n*-butane) and gasoline evaporation (e.g. *i*-pentane) [Choi *et al.*, 2003 and references therein; Table 2]. Tetrachloroethene (C_2Cl_4) is an industrial solvent that is believed to have almost exclusively anthropogenic sources, which makes it a useful tracer for air masses with urban/industrial influence [Atlas *et al.*, 1992; Blake *et al.*, 1997, 2001, 2003a]. Methylene chloride (CH_2Cl_2) is also used as an industrial cleaning solvent [McCulloch *et al.*, 1999; WMO, 1999] and in our experience it often shows a strong correlation with C_2Cl_4 .

Whereas C_2Cl_4 is almost exclusively linked to urban/industrial emissions, most trace gases have multiple sources. For example, although incomplete combustion is the principal global source of ethyne [Seinfeld and Pandis, 1998], the combustion could be in the form of biomass burning, biofuel use, or fossil fuel combustion. In other cases, compounds may have highly unbalanced global budgets with large “missing sources”, as in the case of methyl chloride (CH_3Cl) and methyl bromide (CH_3Br) [Butler *et al.*, 2000]. Although biomass/biofuel burning is the major known source of CH_3Cl , the “missing source” of CH_3Cl is comparable in magnitude to its burning source [Butler *et al.*, 2000]. In these cases, assigning pollution sources to individual principal components requires that multiple species be considered. For example, if ethyne loads together with CH_3Cl then a biomass/biofuel burning source is suspected. This can be further investigated by inspecting whether other species that load onto the same PC are consistent with a burning source. By contrast, a high loading of ethyne with ethene may suggest a fossil fuel combustion source.

The UC-Irvine whole air data set also includes a number of compounds with strong oceanic sources. For example, bromoform (CHBr_3) has a dominant oceanic source and it is often used as a tracer for marine emissions [Class *et al.*, 1986; Atlas *et al.*, 1992; Schneider *et al.*, 1998]. The oceanic source of methyl nitrate (MeONO_2) has been known for over a decade [Walega *et al.*, 1992; Atlas *et al.*, 1993; Chuck *et al.*, 2002], and Blake *et al.* [2003b] provide the strongest evidence yet to establish oceans as the dominant source of MeONO_2 . Dimethyl sulfide (DMS) and methyl iodide (CH_3I) are short-lived compounds with strong marine sources in the equatorial Pacific [e.g. Atlas *et al.*, 1993; Bandy *et al.*, 1996; Blake *et al.*, 1997, 2001; Davis *et al.*, 1999], and CH_3I is considered to be an excellent tracer of marine convection [Davis *et al.*, 1996; Cohan *et al.*, 1999].

Other compounds are not emitted directly from individual emission sources, but are secondary products from the photochemical processing of polluted air masses. For example, whereas oceans are the major source of light ($\text{C}_1\text{-C}_2$) alkyl nitrates, the importance of the marine alkyl nitrate source decreases with increasing carbon number, and photochemical production from the oxidation of parent hydrocarbons is the major source of alkyl nitrates $\geq \text{C}_3$ [Blake *et al.*, 2003b]. Therefore high PC loadings of alkyl nitrates $\geq \text{C}_3$ indicate photochemically processed air.

Whereas the above discussion has linked various tracers to individual emission sources, other tracers can be used to identify individual source regions. For example, although H-2402 and H-1211 are restricted compounds under the Montreal Protocol, their production continues in certain countries because of a lack of substitutes for their critical role as fire extinguishants [Butler *et al.*, 1998; Fraser *et al.*, 1999]. Currently, 90% of the global H-1211 production occurs in the People's Republic of China (PRC), making it an excellent tracer for air that has been influenced by China [Fraser *et al.*, 1999; Blake *et al.*, 2001, 2003a]. As in previous missions,

CH₃Cl also proved to be a useful tracer for air masses with Chinese origin during TRACE-P [Blake *et al.*, 2003a]. Interestingly, CH₃Br was strongly associated with air masses that had passed over Japan and/or South Korea during TRACE-P [Blake *et al.*, 2003a]. The enhanced CH₃Br levels are possibly associated with its use as an import/export crop fumigant in Tokyo, Japan and Inchon, South Korea. Finally, Blake *et al.* [this issue] found substantial carbonyl sulfide (OCS) and carbon disulfide (CS₂) enhancements for a large number of TRACE-P air masses of Chinese and Japanese origin. The major source of CS₂ is chemical industrial emissions (~60%) [Chin and Davis, 1993], and within Asia ~90% of CS₂ emissions are associated with industrial production [Blake *et al.*, this issue]. By comparison, Asian OCS emissions are estimated to be associated primarily with biomass/biofuel combustion (~60%) and industrial emissions (~23%). The vast majority of Asian OCS emissions appear to originate from China, rather than continental Southeast Asia where biomass/biofuel burning is its major source, and may be associated with Chinese coal burning [Blake *et al.*, this issue].

4. Results and Discussion

The TRACE-P whole air data set collected over the western Pacific (east of 112°E, west of 165°W, 0.1-12 km altitude) comprises a matrix of 3458 samples by 47 compounds (Section 2). These data were divided into samples collected north of 25°N (2079 samples) and south of 25°N (1379 samples), and PCA was applied to both data subsets. In previous regional air sampling missions, our group has used 25°N as an arbitrary boundary to separate air with northern and southern hemispheric influences [Blake *et al.*, 2003b]. The TRACE-P data were further separated into three altitudinal bins representing the boundary layer (0.1-2 km), low-to-middle free troposphere (2-7 km), and upper troposphere/lower stratosphere (7-12 km). A Principal Component Analysis was also performed on each of these six data subsets. As a final exercise,

air samples collected at low altitude (0.1-2 km) close to the Asian coast (120-140°E) were also subjected to PCA. Short-lived species (0.5-2 days) such as ethylbenzene and the xylenes were added to this final data set, to determine their contribution to the variability within air masses sampled close to the Asian source region.

4.1 North of 25°N

4.1.1 All Altitudes (0.1-12 km)

The TRACE-P data collected between 25-46°N yield three major PCs and five minor PCs (the top five loadings for the first three PCs are shown in Table 3). The major PCs together represent 77% of the variance in this data set, and the minor PCs each represent 2-3% of the variance. In general the minor PCs were compounds that loaded by themselves, for example CFC-114, DMS and CH₃Br. Components with eigenvalues less than 1.0 account for the remaining variance.

The first PC accounts for one-third of the variance in this data set (Table 3). It shows high loadings of CFC replacement compounds (HCFC-141b, HCFC-142b, HCFC-22, H-134a) and industrial solvents (C₂HCl₃, C₂Cl₄, CH₂Cl₂). Although not shown in Table 3, ethene and the C₄-C₆ alkanes also load onto this PC, with loading values greater than 0.78. These NMHCs reflect the influence of fossil fuel combustion and gasoline evaporation (Section 3.2), and this PC represents a combination of industrial and vehicular sources. The mixture of industrial and vehicular influences is not surprising, given that their emissions are typically co-located in urban centers.

The second PC accounts for one-quarter of the data variance, and it includes high loadings (>0.78) of C₃-C₅ alkyl nitrates and brominated species (CHBr₃, CH₂Br₂). *Blake et al.* [2003b] have shown the importance of photochemical alkyl nitrate production (rather than marine emissions) for alkyl nitrates \geq C₃ (Section 3.2). Thus, the presence of C₃-C₅ (rather than

C₁-C₂) alkyl nitrates indicates photochemically processed air. The presence of brominated species with known oceanic sources most likely suggests the additional influence of marine sources, though the relatively higher loading for CHBr₃ could possibly indicate a minor industrial source (CHBr₃ is used in manufacturing as a cleaning solvent), as speculated upon by *Blake et al.* [2003b]. Other marine tracers such as CH₃I and MeONO₂ also load onto this component, albeit more weakly (0.62 and 0.61, respectively).

The third PC is an interesting mixture of chlorinated species (H-1211, CH₃Cl, 1,2-dichloroethene, CCl₄, CHCl₃), sulfur species (CS₂, OCS), and hydrocarbons (benzene, ethyne). Although the DC-8 and P-3B sampled air from a wide variety of sources during TRACE-P, transport off the Asian continent was frequent [*Fuelberg et al.*, 2003]. The high loadings of H-1211 and CH₃Cl indicate influence from sources within the PRC, including industry and biomass/biofuel burning (Section 3.2). Whereas the major source of CS₂ within Asia is chemical industrial emissions, Asian OCS is associated primarily with emissions from China, from sources including biomass/biofuel combustion and likely coal burning [*Blake et al.*, this issue]. Therefore the third PC represents a mixture of industrial and biomass/biofuel/coal combustion influences with a strong signature from the PRC.

4.1.2 Low Altitudes (0.1-2 km)

The loading table for TRACE-P data collected north of 25°N in the boundary layer (0.1-2 km) is very similar to that for data collected north of 25°N at all altitudes (0.1-12 km). Although the second and third PCs are exchanged and the loading order within each PC varies, the overall results are comparable. This suggests that most of the variance within the larger 0.1-12 km data set is driven by mixing ratio variations in the boundary layer. *Colman et al.* [1998] also found an increased variability for compounds sampled below ~1 km during the Pacific Exploratory

Mission (PEM)-Tropics A, which they attributed to continental outflow from the United States. In the case of TRACE-P, five-day backward trajectories of the sampled air masses indicate that outflow from various regions of Asia was the principal cause of the high variability observed between 0.1-2 km. For example, for transit flights flown north of 30°N, west of 140°E, and below 1.5 km, most of the sampled air passed over Japan and/or other parts of Asia, then descended as it approached the aircraft [Fuelberg *et al.*, 2003]. For TRACE-P flights that deployed from Hong Kong, boundary layer air represented either tropical air in advance of cold fronts, or post-cold-front air originating from central Asia. For flights deployed from Yokota Air Force Base, descending post-frontal continental air was frequently sampled in the boundary layer, sometimes with backward trajectories that passed over highly industrialized regions of Japan and China [Fuelberg *et al.*, 2003].

4.1.3 Middle Altitudes (2-7 km)

In the low-to-middle troposphere (2-7 km) north of 25°N, most of the variability in the whole air data set is explained by compound families that did not feature strongly in the PCA results for 0.1-2 km and 0.1-12 km. Here, the suite of C₃-C₅ alkanes explains one fifth of the variability in the data set, suggesting influence from fossil fuel combustion and gasoline evaporation, and possibly from liquefied petroleum gas (LPG) leakage. Although not shown in Table 3, photochemical oxidation products of the C₃-C₅ *n*-alkanes (namely C₃-C₅ alkyl nitrates) show weak-to-moderate loadings on this component (0.53-0.63), indicating that some of the sampled air masses had undergone photochemical processing.

The second PC accounts for another one fifth of the variability in the data set. Several products consistent with biomass/biofuel burning (CH₃Cl, ethyne, OCS, benzene) are evident. A diverse mixture of other compounds (CCl₄, CH₃I, CHCl₃, EtONO₂) also load onto this

component, reflecting a complex mixture of anthropogenic and possibly natural influences. The presence of short-lived CH_3I (lifetime ~ 4 days) suggests the influence of recent convective uplift (Section 3.2). The complex signals of air masses encountered during TRACE-P has been observed by other researchers. *Streets et al.* [2003] note that in Asia there are millions of individual pollution sources scattered over a very large area. In addition to this complex source signature, *Talbot et al.* [2003] point to extensive frontal uplifting and vertical transport operating over the Asian continent, and they suggest that such continuous intermixing of air parcels of various ages would lead directly to the very complex distributions that were observed for many compounds during TRACE-P.

In contrast to the second principal component, the third PC shows a clear signal, in this case from CFC replacement compounds (Table 3). The high loading of CH_3Br suggests influence from Japan/Korea (Section 3.2). The industrial solvents CH_2Cl_2 and C_2Cl_4 also loaded onto this component (loadings of 0.70 and 0.69, respectively), consistent with the predominantly industrial signature of this component.

4.1.4 High Altitudes (7-12 km)

In the upper troposphere/lower stratosphere (7-12 km) north of 25°N , the first PC is similar to that for 2-7 km, with high loadings of the light alkanes. In the case of 7-12 km, ethane shows a higher loading (0.82) than at 2-7 km (0.61), and *i*-pentane no longer appears in the top five loadings. Whereas the lifetime of *i*-pentane is ~ 5 days in the northern hemisphere (NH) during spring, the lifetime of ethane (~ 2 months in the NH spring) is longer than the ~ 4 week vertical mixing time of the troposphere [*Prather and Jacob*, 1997]. The shift to higher loadings of longer-lived species at higher altitude is consistent with their better mixing (longer reaching influence) throughout the troposphere.

The second PC includes the family of very long-lived halocarbons (CFC-11, CFC-12, CFC-113, CFC-114, CCl_4) as well as comparatively shorter-lived species (H-2402, CH_3CCl_3). Although these compounds are regulated under the Montreal Protocol, their phase-out schedule is more relaxed in developing countries and their usage still continues within Asia. *Blake et al.* [2003a] reported CH_3CCl_3 “hotspots” during TRACE-P in regions downwind of southeast China, Shanghai, Japan and Korea. Using air samples collected along two railroad networks in central and eastern China, *Chan et al.* [this issue] report the continuing use of CFCs in China, particularly in northeast regions. The large-scale variability of the mixing ratio of a trace gas is affected by its sources, sinks and transport [*Johnston et al.*, 2002]. Longer-lived species such as CFCs may be well-mixed (homogeneous) if their residence time in a sampled air mass is long compared to the internal mixing time of the air mass [*Colman et al.*, 2001]. By contrast, the variability of the CFC mixing ratios detected using PCA shows an influence from local sources and is a further indication of their current emission from Asia.

For the third PC, the presence of ethene suggests fossil fuel combustion whereas CH_3Cl suggests biomass/biofuel burning. The high loadings of ethyne and benzene can indicate either fossil fuel combustion or biomass burning (Table 2). Therefore the third PC comprises a suite of species related to incomplete combustion, though we are unable to conclusively determine whether the combustion source is biomass/biofuel burning, fossil fuel combustion, or a mixture of both. Interestingly, CH_3I also loads heavily onto this component. Because the lifetimes of ethene (1-2 days) and CH_3I (~4 days) are much shorter than the vertical tropospheric mixing time (~4 weeks), their contribution to the variability at high altitudes shows the influence of deep convection on the sampled air masses, bringing relatively unprocessed air to the upper troposphere.

4.2 South of 25°N

Although urban/industrial signatures still feature strongly south of 25°N, photochemically processed air masses are more prominent than they were north of 25°N, and clear oceanic signatures also appear (Table 4).

4.2.1 All Altitudes (0.1-12 km)

For whole air samples collected south of 25°N between 0.1-12 km, there were four major PCs and four minor PCs (the first three PCs are shown in Table 4; the fourth major PC, which represents the CFC family, is discussed below). The four major PCs account for 76% of the variance in this data set, and the minor PCs each represent 2-4% of the variance (DMS, H-2402, MeONO₂ and CH₂BrCl each accounted for one of the minor PCs). As was the case north of 25°N, components with eigenvalues less than 1.0 account for the remaining variance.

The first PC accounts for one third of the variance in the data set (Table 4). It comprises C₂-C₅ alkyl nitrates and their longer-lived parent *n*-alkanes (ethane, propane). Oceanic tracers (e.g. CHBr₃, CH₃I, CH₂Br₂) and a variety of continental tracers (e.g. ethyne, OCS, HCFC-142b) also load fairly strongly (0.69-0.77) onto the first PC. The complexity of this signature is reminiscent of PC2 for middle altitudes north of 25°N (Section 4.1.3), although in the present case the high alkyl nitrate loadings show the importance of processed urban air in this more southern data set.

The second PC accounts for one quarter of the variance in the data set. It shows high loadings of tracers related to fossil fuel combustion and gasoline evaporation (ethene, *n*-pentane, *i*-pentane), although the interpretation of PC2 is complicated by the presence of solvents (*n*-hexane, toluene) and CS₂. As discussed above, the major source of Asian CS₂ is chemical

industrial emissions (Section 4.1.1). Therefore this principal component reflects an urban signature, with a mixture of vehicular and industrial influences.

The third PC shows an industrial signature, with weak-to-high loadings of various chlorinated solvents (C_2HCl_3 , C_2Cl_4 , CH_2Cl_2) that in our experience are often enhanced within the same air masses. Methyl chloroform also loads onto PC3, and in previous missions CH_3CCl_3 enhancements have been used to indicate coastal Asian industrial activity [Blake *et al.*, 1996]. However, compared to the lifetimes of the CFCs, the lifetime of CH_3CCl_3 is relatively short (~ 4.9 years; Prinn *et al.*, 2001), and its restriction under the Montreal Protocol has resulted in a sharp decrease in its atmospheric mixing ratio, with TRACE-P mixing ratios only one third those measured in the same region in 1994 [Blake *et al.*, 2003a]. As a result, Blake *et al.* suggest that CH_3CCl_3 is no longer the general indicator of urban air that it was in previous missions. The presence of H-1211 suggests the influence of air masses from the PRC, although we make this association only tentatively because the H-1211 loading is fairly weak.

The fourth major PC (not shown) explains 9% of the variance in this data subset, and the CFC family loads onto this component as follows: CFC-114 (0.82), CFC-113 (0.76), CFC-12 (0.74) and CFC-11 (0.71). In the northern data set, the long-lived CFCs appeared in the top five loadings only for air samples collected in the upper troposphere/lower stratosphere, where the variability of some shorter-lived species may have been dampened during the time required for vertical mixing throughout the troposphere (Section 4.1.4). Here, the variability of the CFC mixing ratios has persisted during transport to the south from Asian source regions.

4.2.2 Low Altitudes (0.1-2 km)

A suite of compounds including the C_2 - C_5 alkyl nitrates explains 28% of the data variance in the boundary layer subset (Table 4), and this PC is similar to PC1 at all southern altitudes (Section

4.2.1). Longer-lived parent hydrocarbons (ethane and propane) also load heavily (0.91 and 0.79, respectively), and several marine species such as MeONO_2 , CHBr_3 and CH_2Br_2 show moderately high loadings (0.68-0.77). These signals indicate photochemically processed air with influence from marine emissions during transport over the western Pacific Ocean. Consistent with this, *Fuelberg et al.* [2003] report that in the lower troposphere during TRACE-P, decaying fronts south of Hong Kong and Japan are a complex region of formerly continental air mixing with more maritime tropical air. The southernmost portions of a few flights deployed from Yokota Air Force Base also sampled pre-frontal air of tropical origin in the lower troposphere, though these encounters were more common for flights deployed from Hong Kong [*Fuelberg et al.*, 2003].

As was the case for PC1, the second PC at low altitudes is almost identical to PC2 at all altitudes for southern latitudes (Section 4.2.1). Similar to our observations north of 25°N , this again suggests that most of the variance within the larger 0.1-12 km data set is driven by mixing ratio variations in the boundary layer. PC2 includes a complex mixture of urban hydrocarbon sources, including fossil fuel combustion (ethene, *i*-butane, *n*-butane), gasoline evaporation (*i*-pentane), and solvent use (*n*-hexane, toluene). Carbonyl disulfide shows the highest loading (0.90), representing industrial emissions from Asia. *Blake et al.* [this issue] estimate that China and Japan account for roughly 45% and 20%, respectively, of industrial CS_2 production within Asia. Halocarbons such as CCl_4 , CH_3Cl and CH_3Br also load onto this component (loadings of 0.64-0.78), suggesting a possible mixed influence from China and Japan/Korea (Section 3.2).

The third PC represents a mixture of industrial halocarbons, including CFCs (CFC-11, CFC-12), CFC replacement compounds (HCFC-22, HFC-134a, HCFC-141b), and industrial degreasing agents and solvents (CH_3CCl_3 , C_2Cl_4 , CH_2Cl_2). Although other halocarbons specific to China (H-1211, CH_3Cl) and Japan/Korea (CH_3Br) do not load onto this component, the high

loadings of CH_3CCl_3 and CFCs are consistent with emissions from China, Japan and Korea (Section 4.1.4). Therefore the third PC represents an Asian industrial signature, with likely source origins in China and Japan/Korea, providing further evidence that at low altitudes the air mass origins were very mixed during TRACE-P.

4.2.3 Mid Altitudes (2-7 km)

The first two PCs in the low-to-mid troposphere (2-7 km) account for 23 and 16% of the data variance, respectively. They are similar to the top two PCs that were obtained between 2-7 km north of 25°N , except with the positions of PC1 and PC2 reversed (Tables 3, 4). South of 25°N , the first PC most likely reflects biomass burning, as indicated by the high loadings of ethyne, CH_3Cl and benzene. The $\text{C}_2\text{-C}_5$ alkyl nitrates also load onto this component, with weak-to-high values (0.59-0.84). The alkyl nitrate loadings increase with decreasing carbon number (increasing lifetime), consistent with aged biomass burning plumes. Southeast Asia was a region of seasonal biomass burning during TRACE-P [Heald *et al.*, 2003], and convective activity over southeast Asia frequently transported boundary layer air to the free troposphere, which was then carried eastward to the sampling region over the Pacific south of 30°N [Kondo *et al.*, this issue]. Nonmethane hydrocarbon emissions from southeast Asia were dominated by biomass burning during TRACE-P [Blake *et al.*, 2003a], consistent with the observed biomass burning signature.

The second PC includes $\text{C}_3\text{-C}_6$ alkanes, indicative of a fossil fuel combustion/gasoline evaporation source. The high loading of short-lived *i*-pentane (~ 5 days) indicates that these air parcels had passed over Asian urban areas fairly recently, as verified from many 5-day backward trajectories of westerly origin at middle altitudes [Fuelberg *et al.*, 2003].

The third PC includes numerous brominated species (CH_2Br_2 , CHBr_3 , CHBr_2Cl , CHBrCl_2). The urban/industrial tracer C_2Cl_4 also loads onto this component, though only weakly

(0.50). Although it only explains 11% of the variance in this data subset, PC3 is the first clear indication of an oceanic suite of gases. *Fuelberg et al.* [2003] note that several backward trajectories for air parcels sampled in the middle troposphere during flights deployed from Hong Kong remain over the Pacific, with origins in the deep tropics and possibly Southern Hemisphere.

4.2.4 High Altitudes (7-12 km)

At the highest altitudes south of 25°N, the first PC explains 37% of the variance and includes a dozen compounds with loadings > 0.86. The general combustion tracers ethyne and benzene both load heavily, which can indicate a fossil fuel source, a biomass/biofuel burning source, or a mixture of both. Because tracers for fossil fuel combustion (*i*-butane, *n*-butane) and gasoline evaporation (*i*-pentane) show stronger loadings (0.89-0.94) than tracers of biomass/biofuel burning (CH₃Cl; loading = 0.65), an urban fossil fuel rather than rural combustion source is suggested. Consistent with this, urban industrial tracers such as H-1211 and C₂Cl₄ also showed strong loadings on this PC (0.83-0.86). Notwithstanding, even though the combustion of biomass/biofuel in rural areas is an important source of CH₃Cl within Asia, comparable CH₃Cl mixing ratios have been measured in both urban and rural areas of China [*T. Wang et al.*, this issue]. Therefore we cannot rule out urban biofuel combustion as a potential contributor to the high loadings of ethyne and benzene.

The second PC is similar to PC2 at high altitudes north of 25°N. The CFC family shows high loadings, together with the long-lived halocarbon CCl₄ (~35 yrs). The relatively shorter-lived compound CH₃CCl₃ (~4.9 yrs) also shows a moderate loading onto this PC (0.60). All six compounds that load onto this PC are restricted under the Montreal Protocol, and their mixing ratio variability is an indication of their continuing use within Asia (Section 4.1.4). As with the

appearance of the CFC family at high altitudes north of 25°N and at low altitudes south of 25°N, increasing vertical or horizontal distance from the Asian source region appears to result in a shift towards higher loadings of longer-lived species.

Only four compounds load onto the third PC, and these represent the suite of oceanic brominated species. Bromoform and CH₂Br₂ both have strong marine sources [Class *et al.*, 1986; Itoh *et al.*, 1994; Yokouchi *et al.*, 1997], and the bromochloromethanes have macroalgae as their main natural source [Class *et al.*, 1986]. We note that unlike CHBrCl₂ and CHBr₂Cl, CH₂BrCl did not load onto this PC. Instead, it loaded onto a minor PC together with ethyl iodide. These results are consistent with early findings from Class *et al.* [1986], whereby CHBrCl₂ and CHBr₂Cl were detected within oceanic surface waters and marine boundary layer air in the Atlantic, but CH₂BrCl was identified only as a minor component emitted from macroalgae. This PC represents the second clear indication of an oceanic suite of gases within the TRACE-P PCA data set.

4.3 Weak or Absent Sources

Whereas PCA is used to identify major pollution sources within large, complex data sets, the weakness or absence of certain emission sources can be of equal interest. Although the DC-8 and P-3B aircraft flew extensively over the western Pacific during TRACE-P, the oceanic tracers were not dominant in the composition of the sampled air masses. This is similar to results we have obtained over even more remote regions of the Pacific during the Biomass Burning and Lightning Experiment (BIBLE-B) [Choi *et al.*, 2003]. Here, the major PCs derived from the TRACE-P data set were predominantly anthropogenic, and they reflect the major impact of Asian pollution sources over the western Pacific.

A signature indicative of liquefied petroleum gas (LPG) was also weak or absent during the TRACE-P mission, in contrast to several Latin American cities where LPG often has a dominant influence on the hydrocarbon composition [Blake and Rowland, 1995; Chen *et al.*, 2001]. We speculate that an LPG signature may begin to emerge from Asia as energy sources within this region change.

4.4 West of 140°E

The above analyses were limited to the 47 compounds that were routinely present at mixing ratios above their detection limit during TRACE-P (Section 2). We also performed PCA on a larger subset of compounds (63 species) that included two additional halocarbons (ethyl chloride, vinyl chloride) and 14 additional NMHCs (propene, acetaldehyde, acetone, butanal, butanone, chlorobenzene, ethylbenzene, hexanal, propanal, methyl propanal, methyl-tertiary-butyl ether (MTBE), *m*-xylene, *o*-xylene, *p*-xylene). Most of these compounds were commonly present at mixing ratios greater than their LOD close to the Asian coast, and the analysis was performed for air samples collected between 30-42°N, 120-140°E and 0.1-2 km. These latitudinal, longitudinal and altitudinal boundaries match those devised by Kita *et al.* [2002]. We have deliberately chosen slightly different latitudinal boundaries than before (30-42°N vs. 25-46°N) to test the robustness of the TRACE-P PCA results.

For this data set there are three major PCs representing 71% of the data variance, and six minor PCs each representing 2-5% of the variance (Table 5). Components with eigenvalues less than 1.0 account for the remaining variance. The top ten loadings are shown in Table 5, rather than just the top five, so that the PCA results can be better compared with those for 25-46°N at low altitudes and at all altitudes (Table 3).

The first major PC includes the same compounds that formed the top five loadings of PC1 for data collected north of 25°N both at low altitudes and at all altitudes (HCFC-141b, C₂HCl₃, C₂Cl₄, HCFC-22, HCFC-142b; Table 3). (We recall that most of the variance within the larger 0.1-12 km data set was found to be driven by mixing ratio variations in the boundary layer; Section 4.2.1). However in the present case, several of the newly-added NMHCs are interspersed in the top 10 loadings, specifically ethylbenzene and the xylene family. These short-lived compounds (0.5-2 days) are emitted from the reforming of petroleum fractions and are used as industrial solvents. As before (Sections 4.1.1; 4.2.1), tracers of fossil fuel combustion and gasoline evaporation also load onto this PC, and it represents a mixture of urban industrial and vehicular sources. In the present case, we are also able to see the influence of short-lived aromatic species on the variability of air masses that were sampled close to the Asian coast.

The second PC is similar to PC2 for data collected at all altitudes north of 25°N, and to PC3 for data collected between 0.1-2 km north of 25°N (Table 3). The C₂-C₅ alkyl nitrates load onto this PC together with ethane, with weaker influences from oceanic tracers such as CHBr₃ and CH₃I. This PC represents photochemically processed air with the likely additional influence from marine sources, or possibly a weak industrial influence (Section 4.1.1). Unlike the case for PC1, no new information was obtained from the addition of the 14 short-lived volatile organic compounds.

The third PC is similar to PC3 for all altitudes north of 25°N, and to PC2 for 0.1-2 km data collected north of 25°N (Table 3). The top five loadings from these latter two data sets also appear in the top 10 loadings shown in Table 5. Chloroform shows a strong loading in Table 5 (0.77), and it likewise had high loadings (>0.71) in the 0.1-2 km and 0.1-12 km data sets

described above (Table 3). Here, benzene and MTBE show weak loadings on this PC, and little additional information was obtained from the analysis of this more comprehensive data set.

Overall, the similarity of the top three PCs in this 63-compound data set compared to the 47-compound data set suggests that the original analysis was adequate for identifying the major modes of composition within the TRACE-P data set and for determining which components drive its variability. In addition, the similar results despite slightly different latitudinal boundaries suggest that the TRACE-P PCA results are fairly robust.

5. Summary and Conclusions

Principal Component Analysis was used to find correlating volatile organic compounds within the TRACE-P whole air data set, which comprises 3458 air samples that were each analyzed for 78 trace gases. Here, a subset of 47 trace gases commonly present at mixing ratios greater than their LOD was used in our analysis (26 halocarbons, 11 nonmethane hydrocarbons, 7 alkyl nitrates, and 3 sulfur species). Knowledge of trace gas emission signatures was used to link each independent group of correlating species (principal component) to individual emission sources. In most cases each principal component was linked to one or two major source influences, though sometimes a principal component revealed a complex mixture of source influences. Other TRACE-P researchers have noted the complexity of air masses that were often sampled during the mission, and have suggested that it is likely caused by millions of individual pollution sources combined with a continuous intermixing of air parcels of various ages.

For air samples collected between 0.1-12 km, 7-46°N, and west of 165°W, industrial sources dominated the composition mode north of 25°N, sometimes with signatures that were unique to China. Tracers of fossil fuel combustion, biomass/biofuel burning, processed urban air, and oceanic emissions were also detected. Using a larger data subset of 63 trace gases that

included 2 additional halocarbons and 14 additional NMHCs, short-lived urban/industrial tracers such as the xylenes were also found to contribute to the measured variability close to the Asian coast. South of 25°N, a complex mixture of processed urban air, air indicating auto exhaust and gasoline evaporation, and oceanic tracers dominated the composition mode. Industrial solvents and degreasing agents also featured strongly, as did the long-lived CFC family.

The variability within the 0.1-12 km data set was driven by trace gas variability within the boundary layer. Both north and south of 25°N, the major principal components for 0.1-12 km were similar to those for 0.1-2 km. These results are consistent with the increased variability for compounds sampled at low altitude, which we have also observed during previous regional airborne studies.

The variability reflected in the PCA loading tables gives an indication of current anthropogenic activities within Asia. Whereas global mixing ratios of short-lived species are spatially variable and will be influenced by local sources upwind of the flight track, longer-lived species may be well-mixed if they are free from the influence of local sources. Here, the variable mixing ratios of long-lived compounds restricted under the Montreal Protocol, including CFCs, CCl₄ and CH₃CCl₃, is an indication of their continuing use within Asia.

The longer-lived species became more dominant in the PCA loading tables for air parcels sampled further from the Asian coast, either at higher altitudes or more southern latitudes. In air sampled north of 25°N for both 2-7 km and 7-12 km altitude bins, the first PC represented fossil fuel emissions and gasoline evaporation. However, in the shift from the free troposphere to the upper troposphere/lower stratosphere, shorter-lived *i*-pentane (~5 days) was replaced by longer-lived ethane (~ 2 mo) in the top five loadings. Similarly, the family of CFCs was identified as a major PC at high altitudes north of 25°N, and at all altitudes south of 25°N. The shift to higher

loadings of longer-lived species with increasing vertical or horizontal distance from the Asian source region is consistent with their better mixing (longer reaching influence) throughout the troposphere.

Principal Component Analysis, coupled with knowledge of trace gas emission signatures, was found to offer a compact, robust method for identifying pollution sources and their major modes of composition within a large regional data set. It could soon play an important role in the analysis of regional whole air data sets as a complement to more familiar methods, first as a starting point for obtaining an overview of these large data sets, and secondly because in PCA calculations all the variables are rendered dimensionless, such that variables with different units (e.g. meteorological variables; aerosol data) can be analyzed together.

Acknowledgments. We dedicate this work to our friend and colleague Murray McEachern. We wish to thank the team at UC-Irvine for their outstanding effort during TRACE-P; the TRACE-P flight crew and ground personnel for their support during the mission; and Mike Cadena, Roy Chesson, and Erica Harper for logistical support. We are grateful to Elliot Atlas and Frank Flocke for their guidance in the development of our alkyl nitrate measurement capability, and to Gao Chen and Martin Schultz for helpful discussions regarding the application of PCA to regional data sets. This research was funded by NASA Grant NCC-1-413.

References

Atlas, E., S. M. Schauffler, J. T. Merrill, C. J. Hahn, B. Ridley, J. Walega, J. Greenberg, L. Heidt, and P. Zimmerman, Alkyl nitrate and selected halocarbon measurements at Mauna Loa Observatory, Hawaii, *J. Geophys. Res.*, 97, D10, 10,331-10,348, 1992.

- Atlas, E., W. Pollock, J. Greenberg, L. Heidt, and A. M. Thompson, Alkyl nitrates, nonmethane hydrocarbons, and halocarbon gases over the equatorial Pacific Ocean during SAGA 3, *J. Geophys. Res.*, 98, 16,933-16,949, 1993.
- Bandy, A. R., D. C. Thornton, B. W. Blomquist, S. Chen, T. P. Wade, J. C. Ianni, G. M. Mitchell, and W. Nadler, Chemistry of dimethyl sulfide in the equatorial Pacific atmosphere, *Geophys. Res. Lett.*, 23, 7, 741-744, 1996.
- Blake, D. R. and F. S. Rowland, Urban leakage of liquefied petroleum gas and its impacts on Mexico City air quality, *Science*, 269, 953-956, 1995.
- Blake, D. R., T.-Y. Chen, T. W. Smith Jr., C. J.-L. Wang, O. W. Wingenter, N. J. Blake, and F. S. Rowland, Three-dimensional distribution of nonmethane hydrocarbons and halocarbons over the northwestern Pacific during the 1991 Pacific Exploratory Mission (PEM-West A), *J. Geophys. Res.*, 101, D1, 1763-1778, 1996.
- Blake, N. J., D. R. Blake, T.-Y. Chen, J. E. Collins Jr., G. W. Sachse, B. E. Anderson, and F. S. Rowland, Distribution and seasonality of selected hydrocarbons and halocarbons over the western Pacific basin during PEM-West A and PEM-West B, *J. Geophys. Res.*, 102, D23, 28,315-28,331, 1997.
- Blake, N. J., D. R. Blake, I. J. Simpson, J. P. Lopez, N. A. C. Johnston, A. L. Swanson, A. S. Katzenstein, S. Meinardi, B. C. Sive, J. J. Colman, E. Atlas, F. Flocke, S. A. Vay, M. A. Avery, and F. S. Rowland, Large-scale latitudinal and vertical distributions of NMHCs and selected halocarbons in the troposphere over the Pacific Ocean during the March-April 1999 Pacific Exploratory Mission (PEM-Tropics B), *J. Geophys. Res.*, 106, 32,627-32,644, 2001.
- Blake, N. J., D. R. Blake, I. J. Simpson, S. Meinardi, A. L. Swanson, J. P. Lopez, A. S. Katzenstein, B. Barletta, T. Shirai, E. Atlas, G. Sachse, M. Avery, S. Vay, H. Fuelberg, C. M. Kiley, and F. S. Rowland, NMHCs and Halocarbons in Asian Continental Outflow during TRACE-P: Comparison to PEM-West B, *J. Geophys. Res.*, 108, D20, 8806, doi:10.1029/2002JD003367, 2003a.
- Blake, N. J., D. R. Blake, A. L. Swanson, E. Atlas, F. Flocke, and F. S. Rowland, Latitudinal, vertical, and seasonal variations of C₁-C₄ alkyl nitrates in the troposphere over the Pacific Ocean during PEM-

- Tropics A and B: Oceanic and continental sources, *J. Geophys. Res.*, 108, D2, doi:10.1029/2001JD001444, 2003b.
- Blake, N. J., Streets, D., J.-H. Woo, I. J. Simpson, J. Green, S. Meinardi, K. Kita, E. Atlas, H. E. Fuelberg, G. Sachse, M. A. Avery, S. Vay, R. W. Talbot, J. E. Dibb, A. R. Bandy, D. C. Thornton, F. S. Rowland, and D. R. Blake, Carbonyl sulfide (OCS) and carbon disulfide (CS₂): Large scale distributions and emissions from Asia during TRACE-P, *J. Geophys. Res.*, this issue.
- Butler, J. H., S. A. Montzka, A. D. Clarke, J. M. Lobert, and J. W. Elkins, Growth and distribution of halons in the atmosphere, *J. Geophys. Res.*, 103, D1, 1503-1511, 1998.
- Butler, J. H., Better budgets for methyl halides?, *Nature*, 403, 260-261, 2000.
- Chan, C. Y., L. Y. Chan, S. C. Zou, I. J. Simpson, K. Y. Chan, and D. R. Blake, Regional mixing ratios, distribution and sources of carbon-containing trace gases in Central and Eastern China, *J. Geophys. Res.*, this issue.
- Chen, T.-Y., I. J. Simpson, D. R. Blake and F. S. Rowland, Impact of the leakage of liquefied petroleum gas (LPG) on Santiago air quality, *Geophys. Res. Lett.*, 28, 2193-2196, 2001.
- Chin, M. and D. D. Davis, Global sources and sinks of OCS and CS₂ and their distributions, *Global Biogeochem. Cycles* 7, 2, 321-337, 1993.
- Choi, Y., S. Elliott, I. J. Simpson, D. R. Blake, J. J. Colman, M. K. Dubey, S. Meinardi, F. S. Rowland, T. Shirai, and F. A. Smith, Survey of BIBLE-B whole air data by Principal Component Analysis (PCA), *J. Geophys. Res.*, doi:10.1029/2002JD002841, 2003.
- Chuck, A. L., S. M. Turner, and P. S. Liss, Direct evidence for a marine source of C₁ and C₂ alkyl nitrates, *Science* 297, 1151-1154, 2002.
- Class, Th., R. Kohnle, and K. Ballschmiter, Chemistry of organic traces (*sic*) in air VII: Bromo- and bromochloromethanes in air over the Atlantic Ocean, *Chemosphere*, 15, 4, 429-436, 1986.
- Cohan, D. S., M. G. Schultz, D. J. Jacob, B. G. Heikes, and D. R. Blake, Convective injection and photochemical decay of peroxides in the tropical upper troposphere: Methyl iodide as a tracer of marine convection, *J. Geophys. Res.*, 104, 5, 5717-5724, 1999.

- Colman, J. J., D. R. Blake, and F. S. Rowland, Atmospheric residence time of CH₃Br estimated from the Junge spatial variability relation, *Science*, 281, 392-396, 1998.
- Colman, J. J., A. L. Swanson, S. Meinardi, B. C. Sive, D. R. Blake, and F. S. Rowland, Description of the analysis of a wide range of volatile organic compounds in whole air samples collected during PEM-Tropics A and B, *Anal. Chem.*, 73, 3723-3731, 2001.
- Davis, D., J. Crawford, S. Liu, S. McKeen, A. Bandy, D. Thornton, F. Rowland, and D. Blake, Potential impact of iodine on tropospheric levels of ozone and other critical oxidants, *J. Geophys. Res.*, 101, 2135-2147, 1996.
- Davis, D., G. Chen, A. Bandy, D. Thornton, F. Eisele, L. Mauldin, D. Tanner, D. Lenschow, H. Fuelberg, B. Huebert, J. Heath, A. Clarke, and D. Blake, Dimethyl sulfide oxidation in the equatorial Pacific: Comparison of model simulations with field observations for DMS, SO₂, H₂SO₄(g), MSA(g), MS, and NSS, *J. Geophys. Res.*, 104, D5, 5765-5784, 1999.
- Edwards, C. H. Jr. and D. E. Penney, *Elementary Linear Algebra*, Prentice Hall, New Jersey, 442 pp., 1988.
- Fraser, P. J., D. E. Oram, C. E. Reeves, S. A. Penkett, and A. McCulloch, Southern hemispheric halon trends (1978-1998) and global halon emissions, *J. Geophys. Res.*, 104, D13, 15,985-16,000, 1999.
- Fuelberg, H. E., C. M. Kiley, J. R. Hannan, D. J. Westberg, M. A. Avery, and R. E. Newell, Meteorological conditions and transport pathways during the Transport and Chemical Evolution over the Pacific (TRACE-P) experiment, *J. Geophys. Res.*, 108, D20, 8782, doi:10.1029/2002JD003092, 2003.
- Heald, C. L., D. J. Jacob, P. I. Palmer, M. J. Evans, G. W. Sachse, H. Singh, and D. Blake, Biomass burning emission inventory with daily resolution: Application to aircraft observations of Asian outflow, *J. Geophys. Res.*, 108, D21, 8811, doi:10.1029/2002JD003082, 2003.
- Henry, R. C. and G. M. Hidy, Multivariate analysis of particulate sulfate and other air quality variables by Principal Components-Part I. Annual data from Los Angeles and New York, *Atmos. Environ.*, 13, 1581-1596, 1979.

- Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2001: The Scientific Basis*, J. T. Houghton *et al.* (eds.), Cambridge University Press, U.K., 2001a.
- Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2001: Mitigation*, J. T. Houghton *et al.* (eds.), Cambridge University Press, U.K., 2001b.
- Itoh, N. and M. Shinya, Seasonal evolution of bromomethanes from coralline algae (Corallinaceae) and its effect on atmospheric ozone, *Marine Chemistry*, 45, 95-103, 1994.
- Jacob, D. J., J. H. Crawford, M. M. Kleb, V. E. Connors, R. J. Bendura, J. L. Raper, G. W. Sachse, J. C. Gille, L. Emmons and C. L. Heald, The transport and chemical evolution over the Pacific (TRACE-P) aircraft mission: Design, execution, and first results, *J. Geophys. Res.*, 108, D20, 9000, doi:10.1029/2002JD003276, 2003.
- Johnston, N. A. C., J. J. Colman, D. R. Blake, M. J. Prather, and F. S. Rowland, On the variability of tropospheric gases: Sampling, loss patterns, and lifetime, *J. Geophys. Res.*, 107, D11, 10.1029/2001JD000669, 2002.
- Kaiser, H. F., The Varimax criterion for analytic rotation in factor analysis, *Psychometrika*, 23, 187-200, 1958.
- Kita, K., Y. Kondo, M. Koike, N. Takegawa, Y. Miyazaki, H. Ikeda, F. Flocke, A. Weinheimer, M. Zondlo, F. Eisele, D. Blake, R. Weber, H. Fuelberg, and G. Sachse, Sources, distribution and partitioning of reactive nitrogen in the lower troposphere over western Pacific during TRACE-P, *EOS Trans. AGU*, 83(47), Fall Meet. Suppl., Abstract A62A-0134, 2002.
- Kondo, Y. *et al.*, Impacts of biomass burning in southeast Asia on ozone and reactive nitrogen over the western Pacific in spring, *J. Geophys. Res.*, this issue.
- McCulloch, A., M. L. Aucott, T. E. Graedel, G. Kleiman, P. M. Midgley, and Y-F. Li, Industrial emissions of trichloroethene, tetrachloroethene, and dichloromethane: Reactive chlorine emissions inventory, *J. Geophys. Res.*, 104, D7, 8417-8427, 1999.
- Paatero, P., Least squares formulation of robust non-negative factor analysis, *Chemometrics and Intelligent Laboratory Systems*, 37, 15-35, 1997.

- Prather, M. J. and D. J. Jacob, A persistent imbalance in HO_x and NO_x photochemistry of the upper troposphere driven by deep tropical convection, *Geophys. Res. Lett.*, 24, N24, 3189-3192, 1997.
- Prinn, R.G., J. Huang, R. F. Weiss, D. M. Cunnold, P. J. Fraser, P. G. Simmonds, A. McCulloch, C. Harth, P. Salameh, S. O'Doherty, R. H. J. Wang, L. Porter, and B. R. Miller, Evidence for substantial variations of atmospheric hydroxyl radicals in the past two decades, *Science*, 292, 1882-1888, 2001.
- Schneider, M., O. Luxenhofer, A. Deissler, and K. Ballschmiter, C₁-C₁₅ alkyl nitrates, benzyl nitrate, and bifunctional nitrates: Measurements in California and South Atlantic air and global comparison using C₂Cl₄ and CHBr₃ as marker molecules, *Environ. Sci. Technol.*, 32, 3055-3062, 1998.
- Seinfeld, J. H. and S. N. Pandis, *Atmospheric Chemistry and Physics*, John Wiley & Sons, New York, 1998.
- Simpson, I. J., N. J. Blake, D. R. Blake, E. Atlas, F. Flocke, J. H. Crawford, H. E. Fuelberg, C. M. Kiley, S. Meinardi, and F. S. Rowland, Production and evolution of selected C₂-C₅ alkyl nitrates in tropospheric air influenced by Asian outflow, *J. Geophys. Res.*, 108, D20, 8808, doi:10.1029/2002JD002830, 2003.
- Statheropoulos, M., N. Vassiliadis, and A. Pappa, Principal component and canonical correlation analysis for examining air pollution and meteorological data, *Atmos. Environ.*, 32, 1087-1095, 1998.
- Streets, D. G., K. Jiang, X. Hu, J. E. Sinton, X.-Q. Zhang, D. Xu, M. Z. Jacobson, and J. E. Hansen, Recent reductions in China's greenhouse gas emissions, *Science*, 294, 1835-1837, 2001.
- Streets, D. G., T. C. Bond, G. R. Carmichael, S. D. Fernandes, Q. Fu, D. He, Z. Klimont, S. M. Nelson, N. Y. Tsai, M. Q. Wang, J.-H. Woo, and K. F. Yarber, An inventory of gaseous and primary aerosol emissions in Asia in the year 2000, *J. Geophys. Res.*, 108, D21, 8809, doi:10.1029/2002JD003093, 2003.
- Talbot R., J. Dibb, E. Scheuer, G. Seid, R. Russo, S. Sandholm, D. Tan, H. Singh, D. Blake, N. Blake, E. Atlas, G. Sachse, and M. Avery, Reactive nitrogen in Asian continental outflow over the western Pacific: Results from the NASA TRACE-P airborne mission, *J. Geophys. Res.*, 108, D21, 8803, doi:10.1029/2002JD003129, 2003.

- Thurston, G. D. and J. D. Spengler, A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston, *Atmos. Environ.*, 19, 9-25, 1985.
- Van de Geer, J. P., *Introduction to Multivariate Analysis for the Social Sciences*, W.H. Freeman, San Francisco, 1971.
- Walega, J. G. *et al.*, Observations of peroxyacetyl nitrate, peroxypropionyl nitrate, methyl nitrate and ozone during the Mauna Loa Observatory Photochemistry Experiment, *J. Geophys. Res.*, 97, 10,311-10,330, 1992.
- Wang, T., C. H. Wong, T. F. Cheung, D. R. Blake, R. Arimoto, K. Baumann, J. Tang, G. A. Ding, X. M. Yu, Y. S. Li, D. D. Streets, and I. J. Simpson, Relationships of trace gases and aerosols and the emission characteristics at Lin'an, a rural site in eastern China during spring 2001, *J. Geophys. Res.*, this issue.
- Wang, Y., C. Shim, N. Blake, D. Blake, Y. Choi, B. Ridley, J. Dibb, A. Wimmers, J. Moodie, F. Flocke, A. Weinheimer, R. Talbot and E. Atlas, Intercontinental transport of pollution manifested in the variability and seasonal trend of springtime O₃ at northern mid and high latitudes, *J. Geophys. Res.*, 2003, in press.
- World Meteorological Organization (WMO), *Scientific Assessment of Ozone Depletion: 1998*, WMO Global Ozone Research and Monitoring Project, Report No. 44, Geneva, 1999.
- Yokouchi, Y., H. Mukai, H. Yamamoto, A. Otsuki, C. Saitoh, and Y. Nojiri, Distribution of methyl iodide, ethyl iodide, bromoform, and dibromomethane over the ocean (east and southeast Asian seas and the western Pacific), *J. Geophys. Res.*, 102, 7, 8805-8809, 1997.
- Yu, T.-Y. and L.-F. W. Chang, Selection of the scenarios of ozone pollution at southern Taiwan area utilizing principal component analysis, *Atmos. Environ.*, 34, 4499-4509, 2000.

Table Captions.

Table 1. Suite of 47 compounds collected in TRACE-P whole air samples and selected for Principal Component Analysis. During TRACE-P these compounds were routinely present at mixing ratios greater than their detection limit.

Table 2. Examples of trace gas emission signatures. See Section 3.2 for references.

Table 3. Principal Component Analysis results for TRACE-P whole air samples collected between 25-46°N and 120°E-165°W. The top five loadings are shown for the first three principal components (PCs), for altitudes between 0.1-12 km (2079 samples, 47 compounds); 0.1-2 km (822 samples); 2-7 km (994 samples); and 7-12 km (263 samples). The first three PCs respectively represent 33, 26 and 17% of the data variance for samples collected between 0.1-12 km; 36, 20 and 18% between 0.1-2 km; 19, 18 and 16% between 2-7 km; and 32, 17 and 12% between 7-12 km.

Table 4. Principal Component Analysis results for TRACE-P whole air samples collected between 7-25°N and 112°E-165°W. The top five loadings are shown for the first three principal components, for altitudes between 0.1-12 km (1379 samples, 47 compounds); 0.1-2 km (413 samples); 2-7 km (716 samples); and 7-12 km (250 samples). The first three PCs respectively represent 34, 24 and 10% of the data variance for samples collected between 0.1-12 km; 28, 26 and 17% between 0.1-2 km; 23, 16 and 11% between 2-7 km; and 37, 10 and 9% between 7-12 km.

Table 5. Principal Component Analysis results for TRACE-P whole air samples collected between 30-42°N, 120-140°E, and 0.1-2 km (279 samples). The top ten loadings are shown for the first three principal components. The first three PCs respectively represent 44, 15 and 12% of the data variance.

Table 1. Suite of 47 compounds collected in TRACE-P whole air samples and selected for Principal Component Analysis. During TRACE-P, these compounds were routinely present at mixing ratios greater than their detection limit.

Halocarbons	Halocarbons Cont'd.	NMHCs	Alkyl Nitrates	Sulfur Compounds
CFC-11	C ₂ Cl ₄	Ethane	MeONO ₂	DMS
CFC-12	C ₂ HCl ₃	Ethene	EtONO ₂	OCS
CFC-113	CHCl ₃	Ethyne	<i>i</i> -PrONO ₂	CS ₂
CFC-114	CHBr ₃	Propane	<i>n</i> -PrONO ₂	
CCl ₄	CH ₂ Cl ₂	<i>i</i> -Butane	2-BuONO ₂	
CH ₃ CCl ₃	CH ₂ Br ₂	<i>n</i> -Butane	2-PeONO ₂	
H-1211	CH ₂ BrCl	<i>i</i> -Pentane	3-PeONO ₂	
H-2402	CHBrCl ₂	<i>n</i> -Pentane		
HCFC-141b	CHBr ₂ Cl	<i>n</i> -Hexane		
HCFC-142b	CH ₃ Cl	Benzene		
HCFC-22	CH ₃ Br	Toluene		
HFC-134a	CH ₃ I			
1,2-DCE	EtI			

DCE = dichloroethene; Me = Methyl; Et = Ethyl; Pr = Propyl; Bu = Butyl; Pe = Pentyl; DMS = dimethyl sulfide.

Table 2. Examples of trace gas emission signatures. See Section 3.2 for references.

Emission source	Tracer compounds
Urban/industry	Tetrachloroethene (C ₂ Cl ₄), Methylene chloride (CH ₂ Cl ₂)
Natural gas	Ethane (C ₂ H ₆), Propane (C ₃ H ₈)
Liquefied Petroleum Gas (LPG)	Propane (C ₃ H ₈), <i>n</i> -Butane (<i>n</i> -C ₄ H ₁₀), <i>i</i> -Butane (<i>i</i> -C ₄ H ₁₀)
Gasoline evaporation	<i>i</i> -Pentane (<i>i</i> -C ₅ H ₁₂)
Fossil fuel combustion	Ethene (C ₂ H ₄), Ethyne (C ₂ H ₂), Benzene (C ₆ H ₆), <i>n</i> -Pentane (<i>n</i> -C ₅ H ₁₂), <i>n</i> -Butane (<i>n</i> -C ₄ H ₁₀), <i>i</i> -Butane (<i>i</i> -C ₄ H ₁₀)
Biomass burning	Ethane (C ₂ H ₆), Ethene (C ₂ H ₄), Ethyne (C ₂ H ₂), Benzene (C ₆ H ₆), Methyl chloride (CH ₃ Cl)
Oceans	Bromoform (CHBr ₃), Methyl nitrate (CH ₃ ONO ₂), Methyl iodide (CH ₃ I), Dimethyl sulfide (DMS)
People's Republic of China	Halon-1211 (H-1211), Carbonyl sulfide (OCS)
Japan/Korea	Methyl bromide (CH ₃ Br)

Table 3.

Loading Order	1 st Principal Component	2 nd Principal Component	3 rd Principal Component
<i>All Altitudes (0.1-12 km)</i>			
1	0.96 (HCFC-141b)	0.92 (2-BuONO ₂)	0.91 (H-1211)
2	0.95 (C ₂ HCl ₃)	0.90 (<i>i</i> -PrONO ₂)	0.85 (CS ₂)
3	0.94 (HCFC-142b)	0.90 (3-PeONO ₂)	0.84 (CH ₃ Cl)
4	0.93 (HCFC-22)	0.88 (CHBr ₃)	0.79 (1,2-DCE)
5	0.93 (C ₂ Cl ₄)	0.88 (<i>n</i> -PrONO ₂)	0.75 (CCl ₄)
<i>Low Altitudes (0.1-2 km)</i>			
1	0.97 (HCFC-141b)	0.92 (H-1211)	0.90 (2-BuONO ₂)
2	0.96 (C ₂ HCl ₃)	0.88 (CCl ₄)	0.89 (3-PeONO ₂)
3	0.95 (C ₂ Cl ₄)	0.86 (CH ₃ Cl)	0.85 (2-PeONO ₂)
4	0.94 (HCFC-22)	0.85 (1,2-DCE)	0.85 (Ethane)
5	0.94 (HCFC-142b)	0.83 (CS ₂)	0.83 (<i>i</i> -PrONO ₂)
<i>Mid- Altitudes (2-7 km)</i>			
1	0.79 (<i>n</i> -Butane)	0.84 (CH ₃ Cl)	0.85 (HCFC-141b)
2	0.78 (<i>n</i> -Pentane)	0.83 (Ethyne)	0.80 (HFC-134a)
3	0.77 (Propane)	0.82 (OCS)	0.78 (HCFC-142b)
4	0.75 (<i>i</i> -Butane)	0.80 (Benzene)	0.73 (HCFC-22)
5	0.71 (<i>i</i> -Pentane)	0.78 (CCl ₄)	0.73 (CH ₃ Br)
<i>High Altitudes (7-12 km)</i>			
1	0.92 (Propane)	0.90 (CFC-113)	0.89 (Ethene)
2	0.91 (<i>n</i> -Butane)	0.89 (CFC-11)	0.82 (Ethyne)
3	0.90 (<i>i</i> -Butane)	0.88 (CCl ₄)	0.80 (CH ₃ Cl)
4	0.82 (Ethane)	0.87 (H-2402)	0.75 (Benzene)
5	0.81 (<i>n</i> -Pentane)	0.83 (CFC-12)	0.65 (CH ₃ I)

Table 4.

Loading Order	1 st Principal Component	2 nd Principal Component	3 rd Principal Component
<i>All Altitudes (0.1-12 km)</i>			
1	0.93 (<i>i</i> -PrONO ₂)	0.90 (Ethene)	0.90 (C ₂ HCl ₃)
2	0.92 (Ethane)	0.89 (CS ₂)	0.74 (C ₂ Cl ₄)
3	0.91 (<i>n</i> -PrONO ₂)	0.88 (<i>n</i> -Pentane)	0.64 (CH ₃ CCl ₃)
4	0.90 (EtONO ₂)	0.88 (<i>n</i> -Hexane)	0.57 (H-1211)
5	0.89 (2-BuONO ₂)	0.87 (<i>i</i> -Pentane)	0.53 (CH ₂ Cl ₂)
<i>Low Altitudes (0.1-2 km)</i>			
1	0.91 (<i>i</i> -PrONO ₂)	0.90 (CS ₂)	0.87 (C ₂ Cl ₄)
2	0.91 (Ethane)	0.87 (Ethene)	0.84 (CFC-11)
3	0.91 (2-BuONO ₂)	0.86 (<i>n</i> -Pentane)	0.83 (HCFC-22)
4	0.90 (<i>n</i> -PrONO ₂)	0.85 (<i>i</i> -Pentane)	0.80 (HFC-134a)
5	0.88 (3-PeONO ₂)	0.84 (<i>n</i> -Hexane)	0.69 (CH ₃ CCl ₃)
<i>Mid-Altitudes (2-7 km)</i>			
1	0.91 (Ethyne)	0.96 (<i>n</i> -Butane)	0.80 (CH ₂ Br ₂)
2	0.88 (CH ₃ Cl)	0.96 (<i>n</i> -Pentane)	0.77 (CHBr ₃)
3	0.88 (Benzene)	0.92 (<i>i</i> -Butane)	0.73 (CHBr ₂ Cl)
4	0.84 (EtONO ₂)	0.89 (Propane)	0.65 (CHBrCl ₂)
5	0.80 (CH ₃ Br)	0.87 (<i>i</i> -Pentane)	0.50 (C ₂ Cl ₄)
<i>High Altitudes (7-12 km)</i>			
1	0.94 (Ethyne)	0.88 (CFC-11)	0.87 (CHBr ₂ Cl)
2	0.94 (<i>n</i> -Butane)	0.82 (CCl ₄)	0.82 (CH ₂ Br ₂)
3	0.92 (<i>i</i> -Butane)	0.81 (CFC-12)	0.82 (CHBr ₃)
4	0.89 (<i>i</i> -Pentane)	0.80 (CFC-113)	0.71 (CHBrCl ₂)
5	0.89 (Benzene)	0.73 (CFC-114)	-

Table 5.

Loading Order	1 st Principal Component	2 nd Principal Component	3 rd Principal Component
<i>Low Altitudes (0.1-2 km)</i>			
1	0.98 (HCFC-141b)	0.94 (2-BuONO ₂)	0.90 (Chlorobenzene)
2	0.97 (Ethylbenzene)	0.93 (3-PeONO ₂)	0.88 (H-1211)
3	0.97 (C ₂ HCl ₃)	0.91 (2-PeONO ₂)	0.84 (CH ₃ Cl)
4	0.97 (<i>m</i> -xylene)	0.88 (<i>i</i> -PrONO ₂)	0.81 (CCl ₄)
5	0.97 (<i>o</i> -xylene)	0.84 (<i>n</i> -PrONO ₂)	0.77 (CHCl ₃)
6	0.97 (<i>p</i> -xylene)	0.74 (Ethane)	0.70 (CS ₂)
7	0.96 (HCFC-142b)	0.74 (EtONO ₂)	0.70 (1,2-DCE)
8	0.96 (C ₂ Cl ₄)	0.65 (CHBr ₃)	0.57 (Benzene)
9	0.96 (HCFC-22)	0.58 (CH ₃ I)	0.54 (MTBE)
10	0.96 (<i>n</i> -Hexane)	0.55 (Ethyne)	-